

Databases and ontologies

QuickGO: a web-based tool for Gene Ontology searching

David Binns*, Emily Dimmer, Rachael Huntley, Daniel Barrell, Claire O'Donovan and Rolf Apweiler

European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Received on July 13, 2009; revised on September 4, 2009; accepted on September 7, 2009

Advance Access publication September 10, 2009

Associate Editor: Dmitrij Frishman

ABSTRACT

Summary: QuickGO is a web-based tool that allows easy browsing of the Gene Ontology (GO) and all associated electronic and manual GO annotations provided by the GO Consortium annotation groups. QuickGO has been a popular GO browser for many years, but after a recent redevelopment it is now able to offer a greater range of facilities including bulk downloads of GO annotation data which can be extensively filtered by a range of different parameters and GO slim set generation.

Availability and Implementation: QuickGO has implemented in JavaScript, Ajax and HTML, with all major browsers supported. It can be queried online at <http://www.ebi.ac.uk/QuickGO>. The software for QuickGO is freely available under the Apache 2 licence and can be downloaded from <http://www.ebi.ac.uk/QuickGO/installation.html>

Contact: goa@ebi.ac.uk; dbinns@ebi.ac.uk

1 INTRODUCTION

The Gene Ontology (GO) Consortium (<http://www.geneontology.org/>) provides three detailed ontologies of terms describing the molecular functions that gene products normally perform, the biological processes they are involved in and the subcellular locations (cellular components) where they are located. Associations between gene products and these terms are assigned by many biological databases providing detailed functional descriptions. A single gene product may be annotated to multiple GO terms, detailing a range of functional attributes, using both manual and electronic annotation methods (distinguished by evidence code), with references from PubMed or for an electronic method. GO has proved particularly useful in providing biological insights for large proteomic or genomic datasets (Dimmer *et al.*, 2008).

The Gene Ontology Annotation (GOA) project (<http://www.ebi.ac.uk/goa>) provides high-quality electronic and manual annotations of GO terms to UniProt Knowledgebase (UniProtKB) entries (Barrell *et al.*, 2009; UniProt Consortium, 2009). Annotations created by GOA are collated with those from over 20 external databases (including GO Consortium members, e.g. MGI and FlyBase, and specialist databases, e.g. Human Protein Atlas and IntAct). GOA's combined activities provide an extensive, publicly available GO resource containing more than 45 million annotations for over 6 million proteins in 1 97 000 taxa (July 2009 release).

The QuickGO browser (<http://www.ebi.ac.uk/QuickGO>) provides user with access to all GO term information and the GO annotations

released by GOA. It has recently been redeveloped to offer users a number of unique features, including the ability to tailor annotation sets using multiple filtering options as well as to construct subsets of the GO (GO slims) to map-up annotations allowing a general overview of the attributes of a set of proteins.

2 OVERVIEW OF Quick GO

The QuickGO interface has been built using JavaScript, Ajax and HTML, and uses an optimized and streamlined set of files to facilitate large queries and data extraction. Besides providing GO term and annotation data, QuickGO also imports protein descriptions and taxonomic hierarchies from UniProtKB. The tool is updated with GO term information nightly and GO annotations weekly, ensuring that it is one of the most up-to-date GO resources.

2.1 Simple searching

QuickGO users can view and search information provided for GO terms (identifiers, words/phrases in the title or definition, cross-references and synonyms), as well as protein data from UniProtKB (accession numbers, names and gene symbols). Results are ranked so that terms most closely matching the query are returned first. Individual words and combinations of words are scored according to the field in which they occur and their frequency within GO.

When a GO term is selected, QuickGO presents a tabbed view of the term's information (including definition, cross-references and child terms). Interactive graph and table views of the term's position in the GO as well as a page displaying all annotations to the selected term are presented. Statistics for each GO term are available, showing which terms are frequently concurrently annotated, aiding curators to annotate consistently.

QuickGO is updated weekly with protein names, gene symbols, accessions and taxonomy data from UniProtKB. Single or multiple protein accessions can be queried and selected proteins will display all associated GO annotations, both electronic and manual.

2.2 Annotation download facility

The GOA annotation set is a central component of QuickGO that can be filtered using multiple combinations of parameters. Tailoring annotation sets has become increasingly necessary as the number of GO annotations to new sequences has increased rapidly, causing users to be overwhelmed when downloading and filtering the GOA- UniProt gene association file (2.2 GB in July 2009).

*To whom correspondence should be addressed.

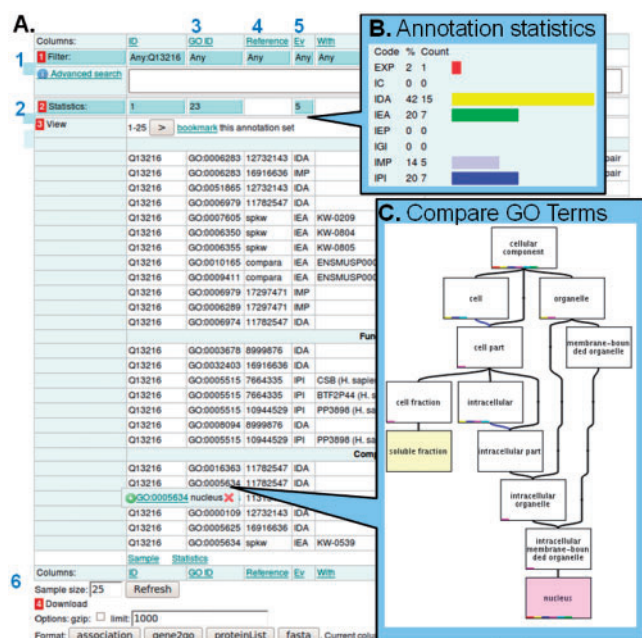


Fig. 1. (A) Protein Annotation view. (1) Filtering options. (2) Statistics. The table contains all the information for a protein-GO term association, including: (3) names and identifiers of GO terms; (4) reference for the annotation; (5) GO evidence code for the annotation. (6) Annotation sets can be downloaded in various formats. (B) Statistics for the 'Evidence' column showing the percentage and count of each evidence code for this set. (C) Nucleus term (purple) compared with another GO term (yellow).

In total, nine different parameters are available for users to create their desired subset of annotations. Filtering options include:

Specific protein(s): users can submit lists of identifiers and map between identifier types (e.g. UniProtKB, Ensembl, RefSeq, MGI and FlyBase identifiers). Over 14 million identifiers are available.

Evidence codes: annotations can be selected for specified evidence codes, for example, representing results from electronic (IEA) or experimental methods (IMP, IGI, IPI, IDA, IEP and EXP).

Qualifier data: qualified annotations can be filtered, for example, negative ('NOT') annotations can be removed.

Taxonomic data: As QuickGO uses the UniProtKB taxonomic hierarchy, it can provide annotation sets for large taxonomic groups, e.g. all flowering plants, or for a single species.

GO terms: annotations using selected GO terms can be viewed (see also Section 2.3 for GO slim).

More complex queries are made using the advanced search allowing users to apply a combination of Boolean operators (AND, NOT, OR) to their queries.

Once an annotation set has been selected, QuickGO provides a paginated view of annotations and statistics. Statistics are calculated on-the-fly, and provide detailed summaries of, e.g. the distribution of GO terms or evidence codes present in the specified dataset (Fig. 1B). Much of the data in the annotation table is clickable (including GO terms, taxonomy information, identifiers) to provide further information or links to relevant web pages. Finally, users

can specify the format of their downloaded data; either in gene association, FASTA (sequences) or gene2go (NCBI standard for annotations) formats or alternatively as a list of protein/gene identifiers (Fig. 1A).

2.3 GO slim generation and annotation mapping

QuickGO not only filters annotations to a specific set of GO terms, but also allows the reaction or adaption of GO slims. Users can collect a set of GO terms, view them as a graph and use them to construct a GO slim. Annotations can be 'mapped up' or 'slimmed' to these selected GO terms. Predetermined GO slims created by groups in the GO Consortium are also provided. These combined facilities allow users to extensively tailor both the GO slim sets and the associated GO annotations (<http://www.ebi.ac.uk/QuickGO/GMultiTerm>).

2.4 Web services

All data provided by QuickGO can be queried remotely, both for GO term information and annotation data. These web services are fully integrated, so that the filtering options and datasets available are fully synchronized between the browsable and web service interfaces. The web services have been designed for ease of use, providing a REST-style query interface with results displayed in standard compliant gene association, OBO or XML formats.

3 CONCLUSION AND PERSPECTIVE

QuickGO includes documentation to support all levels of user, including web tutorials (www.ebi.ac.uk/QuickGO/tutorial.html) detailing the main aspects of the tool. Each QuickGO page (including filtered views of annotation sets) can be bookmarked to enable users to revisit specific annotation sets.

QuickGO is the most comprehensive source of GO annotation data, other GO browsers, such as AmiGO (<http://amigo.geneontology.org>), are only able to supply manual or species-specific annotations. No other GO browser currently provides such extensive identifier mapping or taxonomic querying facilities.

ACKNOWLEDGEMENTS

We would like to thank Manjula Thimma for programming, and Ruth Lovering and Varsha Khodiyar for testing the tool.

Funding: Biotechnology and Biological Sciences Research Council, Tools and Resources Fund (grant number BB/E023541/1).

Conflict of Interest: none declared.

REFERENCES

- Dimmer, E.C. et al. (2008) The gene ontology - providing a functional role in proteomic studies. *Proteomics*, [Epub ahead of print, doi:10.1002/pmic.200800002].
- Barrell, D. et al. (2009) The GOA database in 2009—an integrated gene ontology annotation resource. *Nucleic Acids Res.*, **37**, D396–D403.
- UniProt Consortium (2009) The Universal Protein resource (UniProt). *Nucleic Acids Res.*, **37**, D169–D174.